ENCODE 2020: From Elements to Function

ENCODE PIs' Vision for Functional Genomics

Joe Ecker
Salk/HHMI

on behalf of
Brad B, David G, Mike S, John S, Barbara W

Core accomplishments of the ENCODE project:

- Creation of vast, accessible catalogs of regulatory DNA, transcription factor occupancy and histone modification patterns, and RNA transcripts, as well as a standard curation of protein-coding and non-coding genes (GENCODE).

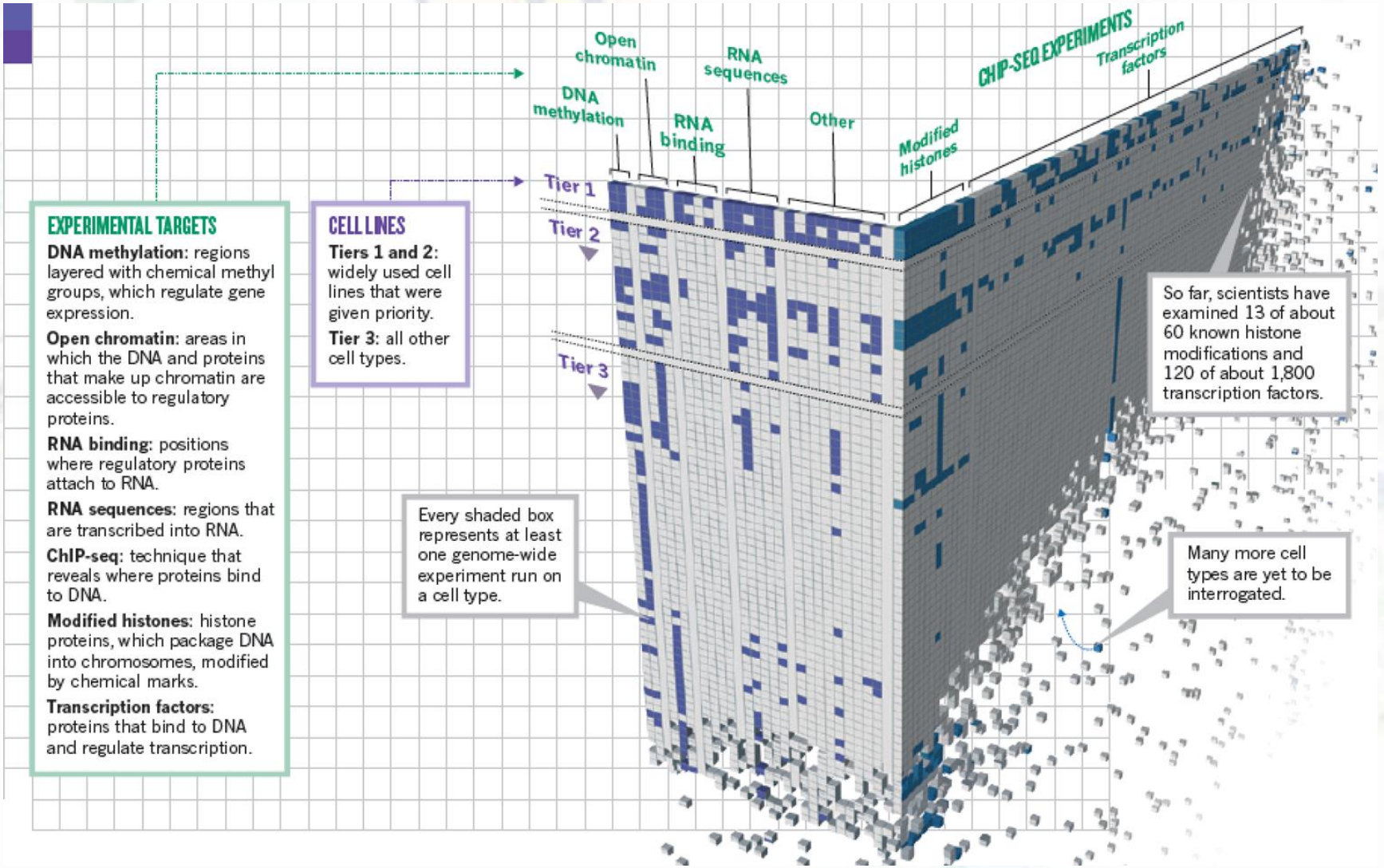
- Development and dissemination of standards and experimental methods for producing high-quality, reproducible data in a cost efficient manner from major assay types including ChIP-seq, RNA-seq, and DNase-seq.

- Development and dissemination of algorithms and software for analysis of major regulatory genomic data types, as well as tools and methods for integrating functional genomic data.

- ENCODE has trained a new generation of fellows and students in genome science, who continue to play major roles in methods development, data generation and analysis.

Encyclopedia of DNA Elements

-the project is still far from complete-



A Catalog of DNA Elements

[1]catalog

noun | cat·a·log | \'ka-tə-ˌlȯg, -ˌläg\

Share | +1 | Tweet

: a book containing a list of things that you can buy, use, etc., and often pictures

: a group of similar or related things

Imminent challenges and the role of ENCODE

- Despite rapid progress across the field of functional genomics, identifying all functional elements of the human genome is an unfulfilled aspiration

- ENCODE data reveal greater diversity (combinatorial activation patterns and modification signatures) and greater numbers (up to millions) of elements than anticipated.

- The next phase must leverage and integrate emerging technological, computational and biological strategies to tackle complex biological problems such as cell differentiation and the etiology of disease.

- High-throughput approaches for mapping genomic features (biochemical and otherwise) will be complemented by new tools for high-throughput genome engineering and systematic functional perturbation.

Function

- Layer 1: Completing the catalog of elements
- Layer 2: Connecting elements with their cognate gene(s)
- Layer 3: Transforming the catalog of elements into a full-fledged encyclopedia
- Layer 4: From general to specific: individual variation in sequence elements and its impact on quantitative phenotypes and disease

Layer 1: Completing the Catalog of Elements

- **New cell and tissue types.**

The human body comprises over 400 recognized cell types based on classical microscopic and histochemical modes of analysis; the true number is likely higher.

- **New types of elements**.

The genome encodes diverse functional and physical interactions that are poorly understood (e.g., with 1000s of regulatory factors that bind DNA or RNA)

- **Condition-specific elements**.

Many elements are activated in response to particular external stimuli (e.g., steroid response elements) or intrinsic programs such as differentiation.

Layer 1: Completing the Catalog of Elements will require:

- A new generation of mapping/discovery tools
- Substantially higher sample throughput (>10X over current platforms), while maintaining high cost efficiency
- Routine application to small numbers of cells (500-50,000 cell range) to enable penetration of diverse biologically meaningful compartments
- Critically, the above must be achieved without erosion in resolution or data quality compared with current gold-standard assays.

Layer 1: Completing the Catalog of Elements Enhancements:

- Create a community-focused data coordination center to augment and expand consortium efforts by assembling, curating and making easily publicly accessible the high-quality data and corresponding metadata generated by diverse expert community investigators.

- Create a truly global resource by systematically integrating data from other large-scale functional genomics projects (e.g., GGR, GTEx, Epigenomics Consortia) with ENCODE and community data into an easily accessible comprehensive reference.

- The above efforts have the potential to make ENCODE data – and those of many other projects ranging from focused R01s to large consortia – more universal, accessible, and useful.

Layer 2: Connecting elements with their cognate gene(s)

- Local chromatin interactions reveal enhancer/promoter interactions
- Genome-wide survey of long-range chromatin interactions
- Functional analysis of long-range regulatory elements

Layer 2: Connecting elements with their cognate gene(s)

Approaches:

- Activity correlation
- Physical interaction
- Knockouts.

Layer 2: Connecting elements with their cognate gene(s)

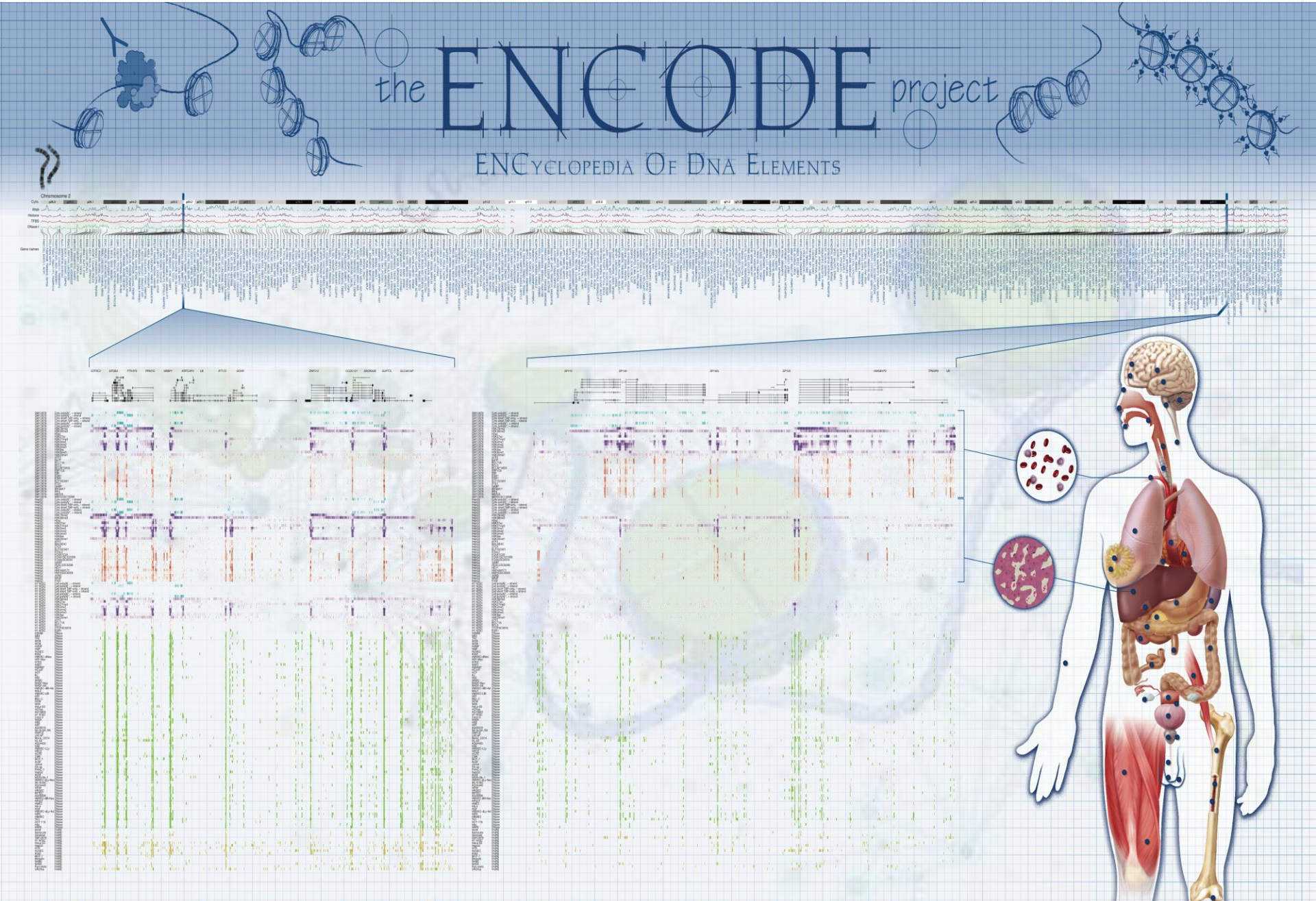
Approaches:

- **Activity correlation**

Element biochemical signatures are tightly correlated with the appearance of activating features at the promoters of their cognate gene(s)

Elements show cell selectivity, analysis of these co-activation patterns over dozens or even hundreds of cell types can systematically connect elements with target genes.

Activity correlation



Layer 2: Connecting elements with their cognate gene(s)

Approaches:

- Physical interaction

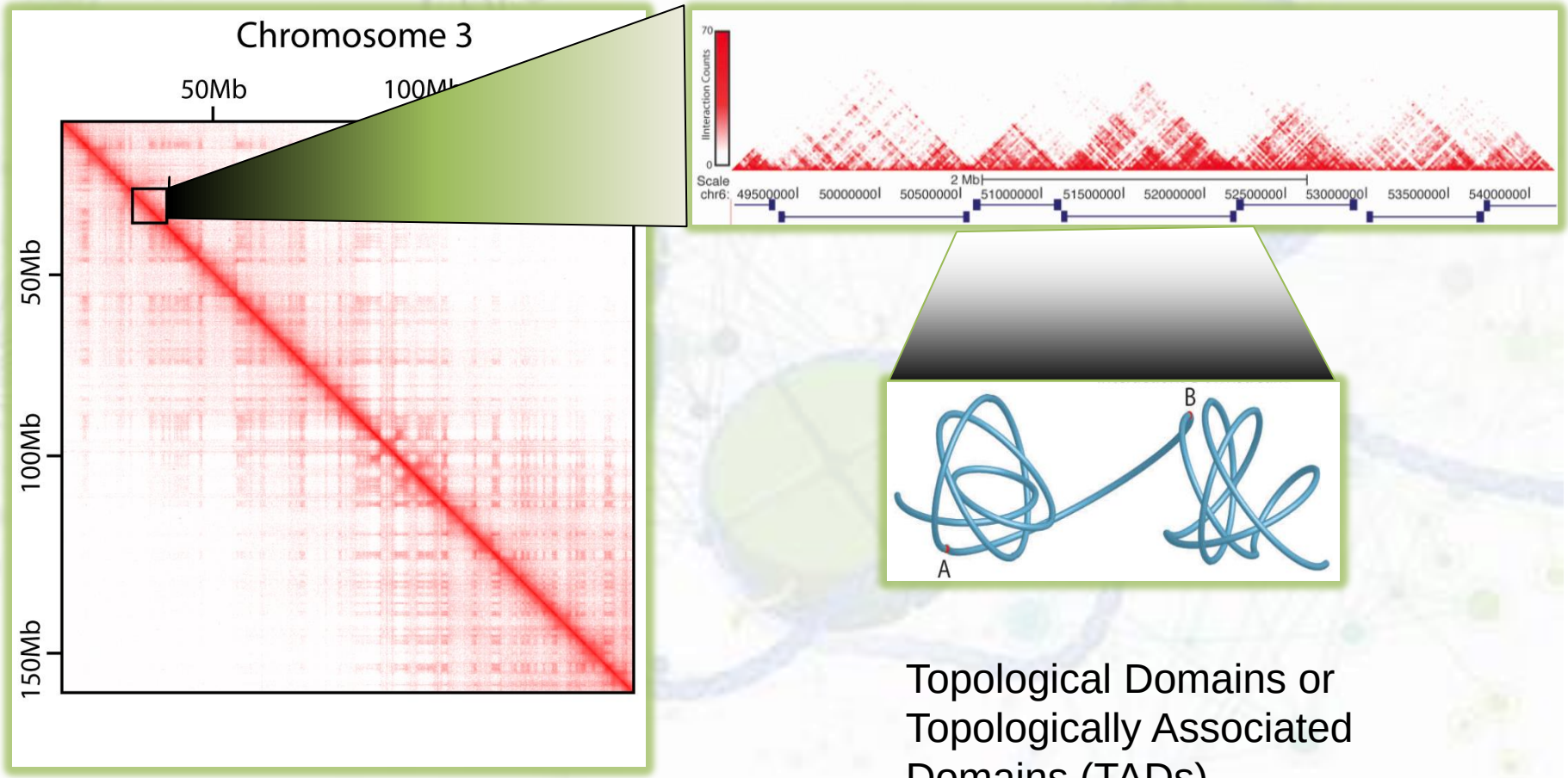
Distal element contacts with their target promoters (or other elements) can now be routinely measured with several experimental strategies (e.g., 5C, HiC, ChIA-PET etc.).

Understanding of how such interactions – or which interactions – are most significant from the functional perspective is still nascent.

- .

Physical interaction

Genome-wide analysis of higher order chromatin structure in human and mouse cells



Topological Domains or Topologically Associated Domains (TADs)

Higher Hi-C frequency = shorter spatial distance

Lower Hi-C frequency = longer spatial distance

Layer 2: Connecting elements with their cognate gene(s)

Approaches:

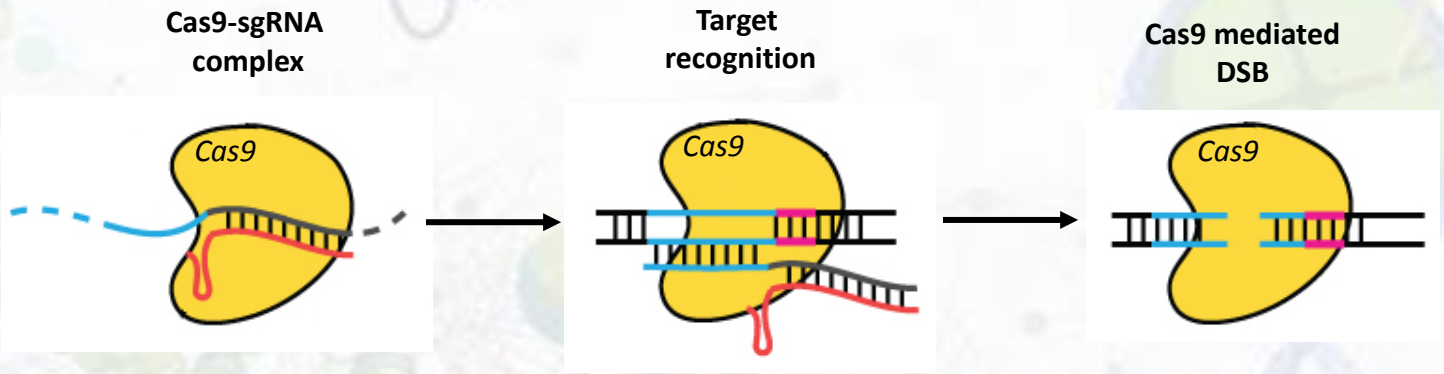
- Knockouts.

Reverse genetics in an isogenic setting is a powerful approach for establishing both function per se, and specific connections between a given DNA segment and control of specific genes.

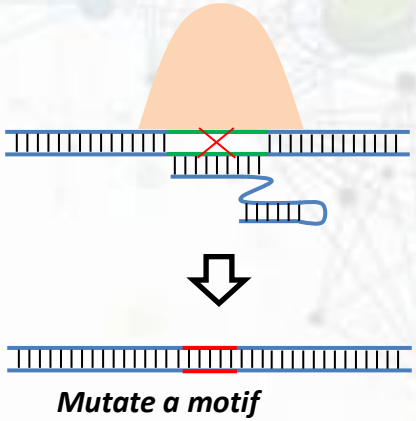
Knockouts

Using CRISPR/Cas9 to mutate enhancers

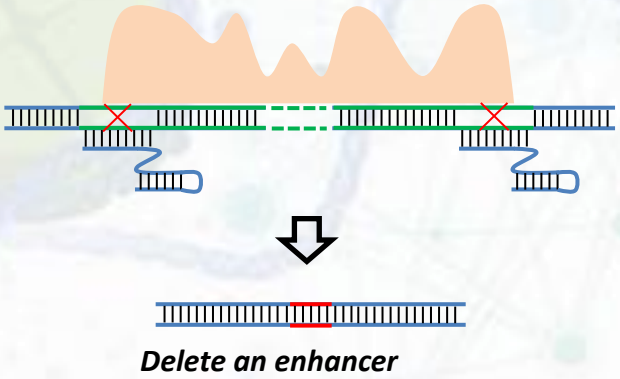
a



b

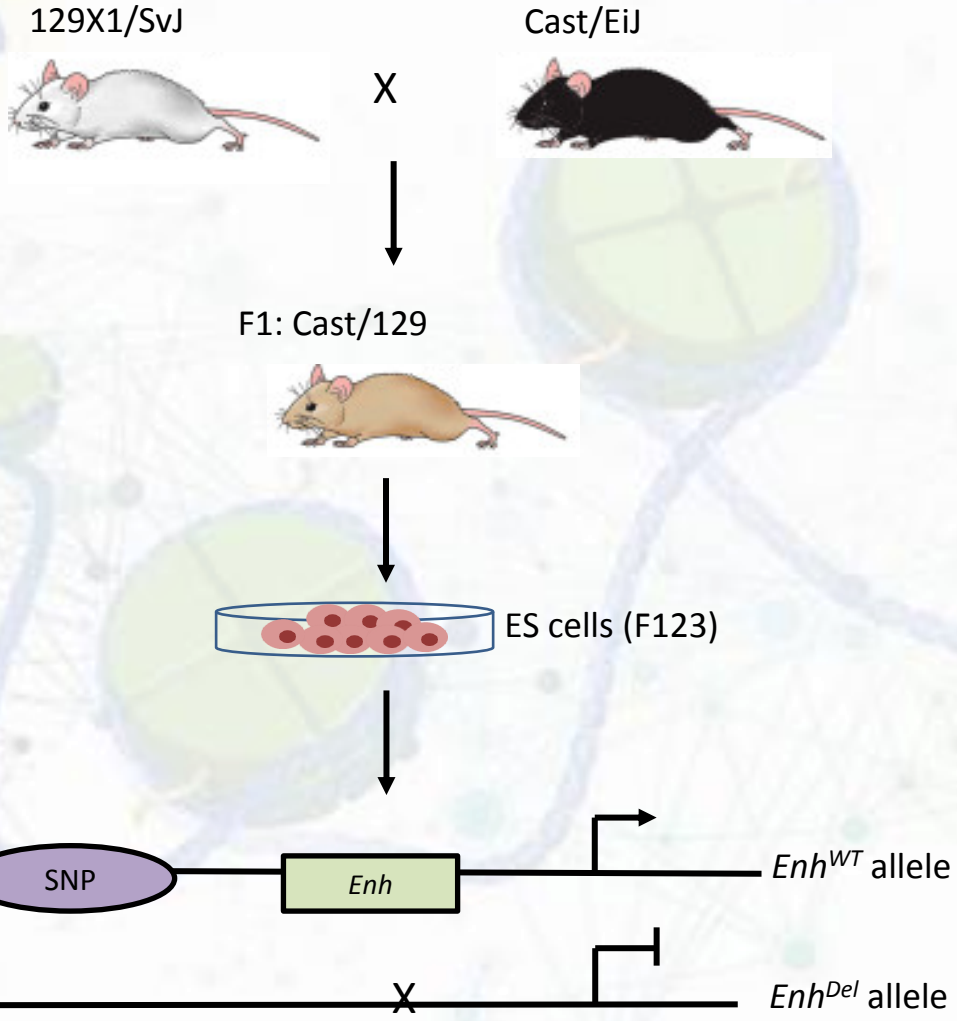


c



Validate the cis-regulatory functions of enhancers

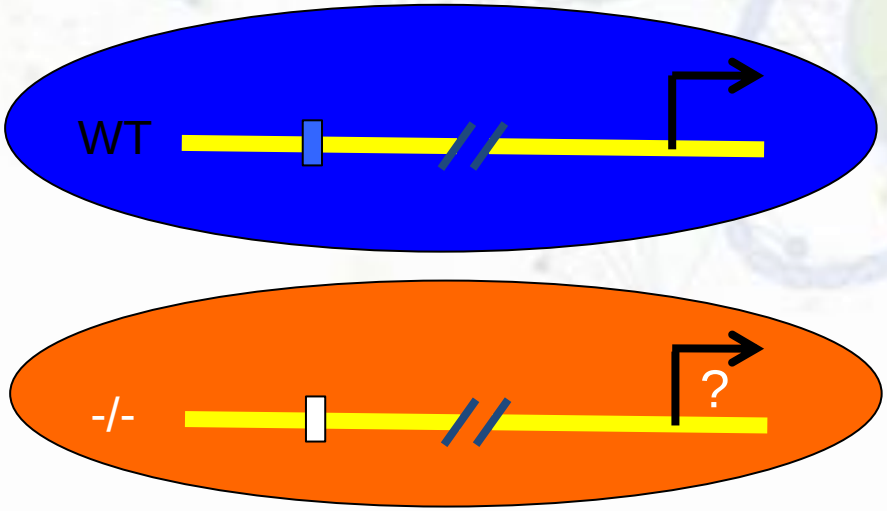
- Enhancer knockout provide direct evidence
  - Test the transcription enhancing effect
  - Test if the effect is in *cis*.

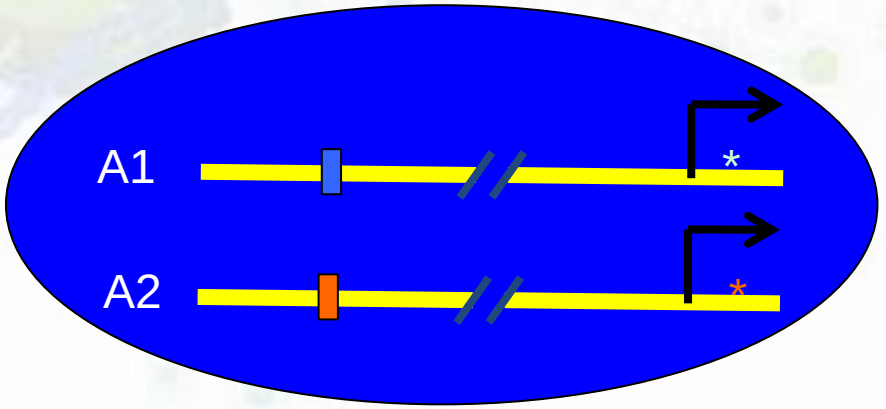


Strategies for functional study of enhancers

- Introduce mutations into each enhancer in their endogenous locus and test for changes in gene expression
  - Pros: most direct
  - Cons: low throughput; may not applicable to humans

- Exploit the naturally occurring sequence variants (SNPs) between the two copies of DNA in each cell
  - Pros: global and genome-wide
  - Cons: need to know the haplotypes





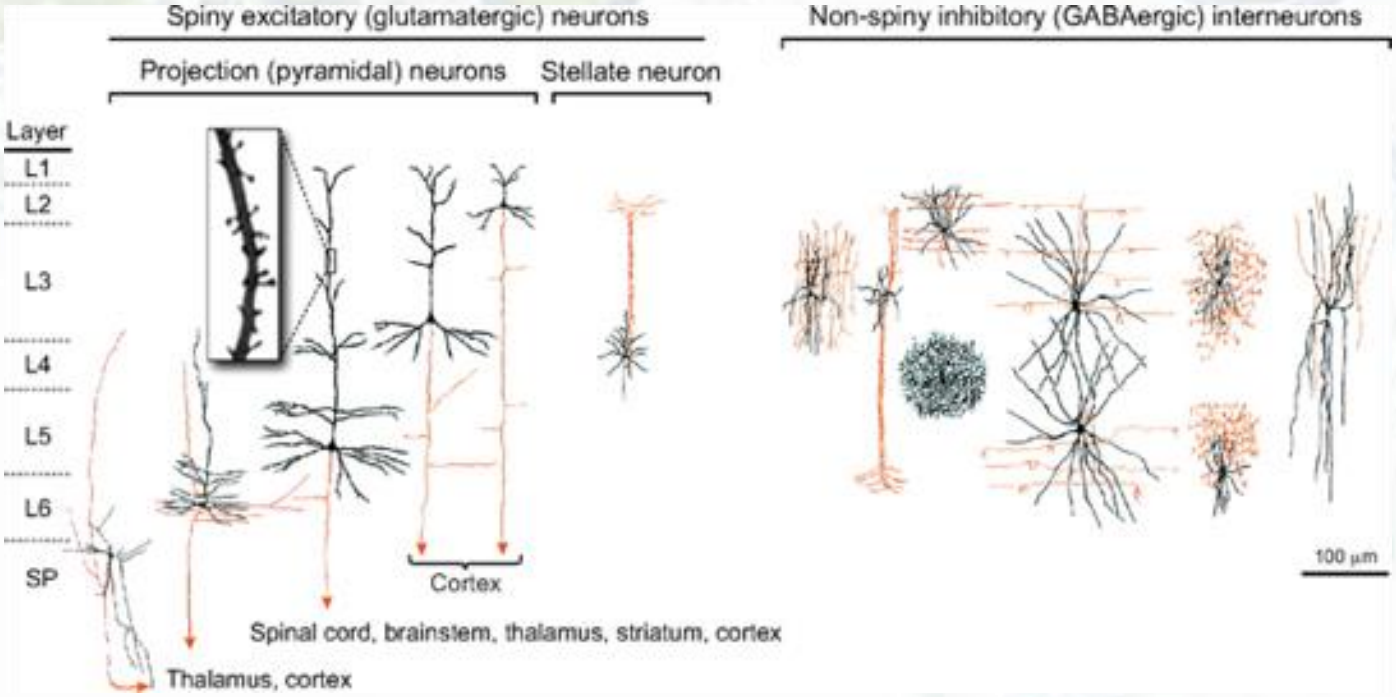
Layer 2: Connecting Elements with their cognate gene(s)

Substantial challenges exist:

- Different categories of elements will impact different features – from transcription initiation to elongation to splicing to local and regional chromatin states – many of which may not be readily detectable with conventional assays.
- Cellular and genomic context sensitivity is likely to be the rule – individual elements have evolved within a specific chromatin context, and at specific distances from genes and other nearby elements.
- Many elements are ‘primed’ or ‘memory’ sites – elements that are detectable biochemically (e.g., paused RNA transcripts, certain histone modifications or hypersensitivity) yet impotent within a particular context in which additional activating signals are missing.

Layer 2: Connecting Elements with their cognate gene(s)

- **Cellular and genomic context sensitivity is likely to be the rule** – individual elements have evolved within a specific chromatin context, and at specific distances from genes and other nearby elements.



Layer 2: Connecting Elements with their cognate gene(s)

- Many elements are ‘primed’ or ‘memory’ sites – elements that are detectable biochemically (e.g., paused RNA transcripts, certain histone modifications or hypersensitivity) yet impotent within a particular context in which additional activating signals are missing.

Layer 2: Connecting elements with their cognate gene(s) will require

- Development of novel genome-scale assays
- Systematic experimental perturbations
- Integrative computational analysis

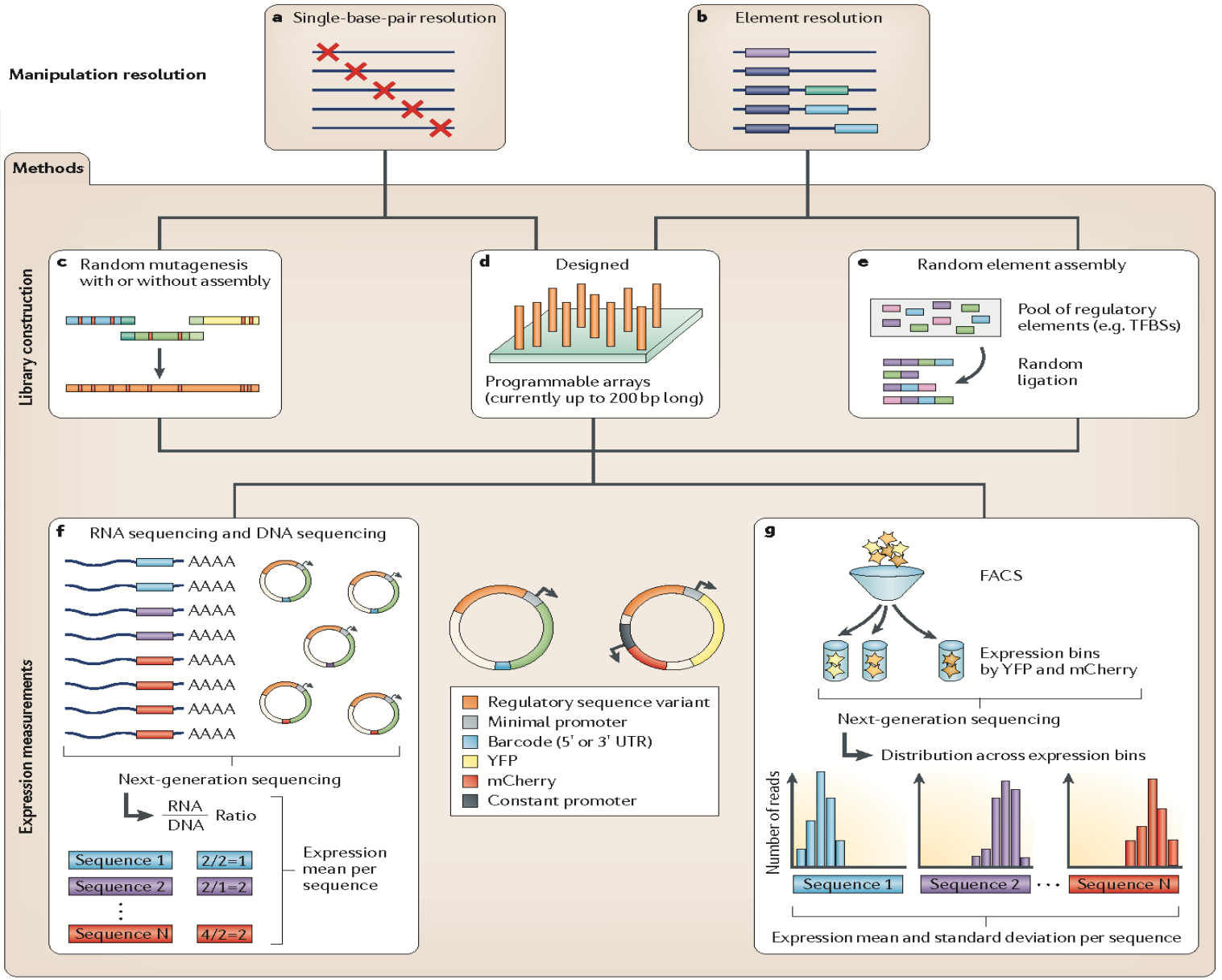
Above requirements will challenge the limits of high-throughput functional genomics platforms

This type of effort is well suited to a consortium approach, and the nature of the resulting data will be of immediate and ongoing utility for the community

Layer 3:Transforming the Catalog of Elements into a full-fledged Encyclopedia

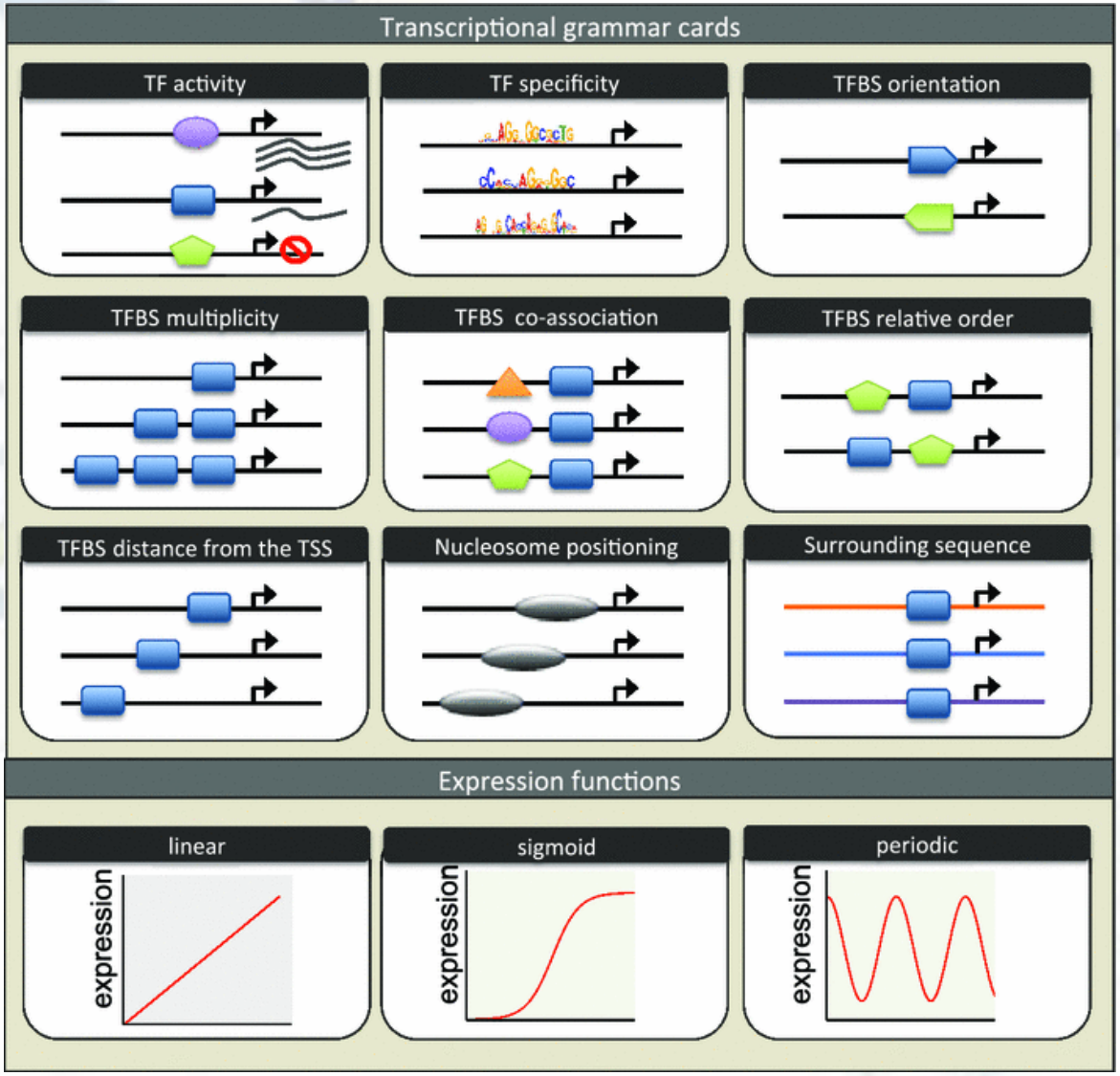
- Transforming the catalog into a full-fledged encyclopedia will require systematic categorization functional elements.
- Categorizing sequence elements into functional behavioral classes
- Not only where are the element but what and how
- Identify all of the major categories of functional elements encoded by the genome

Dissection of regulatory sequences using massively parallel reporter assays



Levo and Segal (2014) NATURE REVIEWS | **GENETICS**

Understanding the Grammar of Gene Expression Regulation



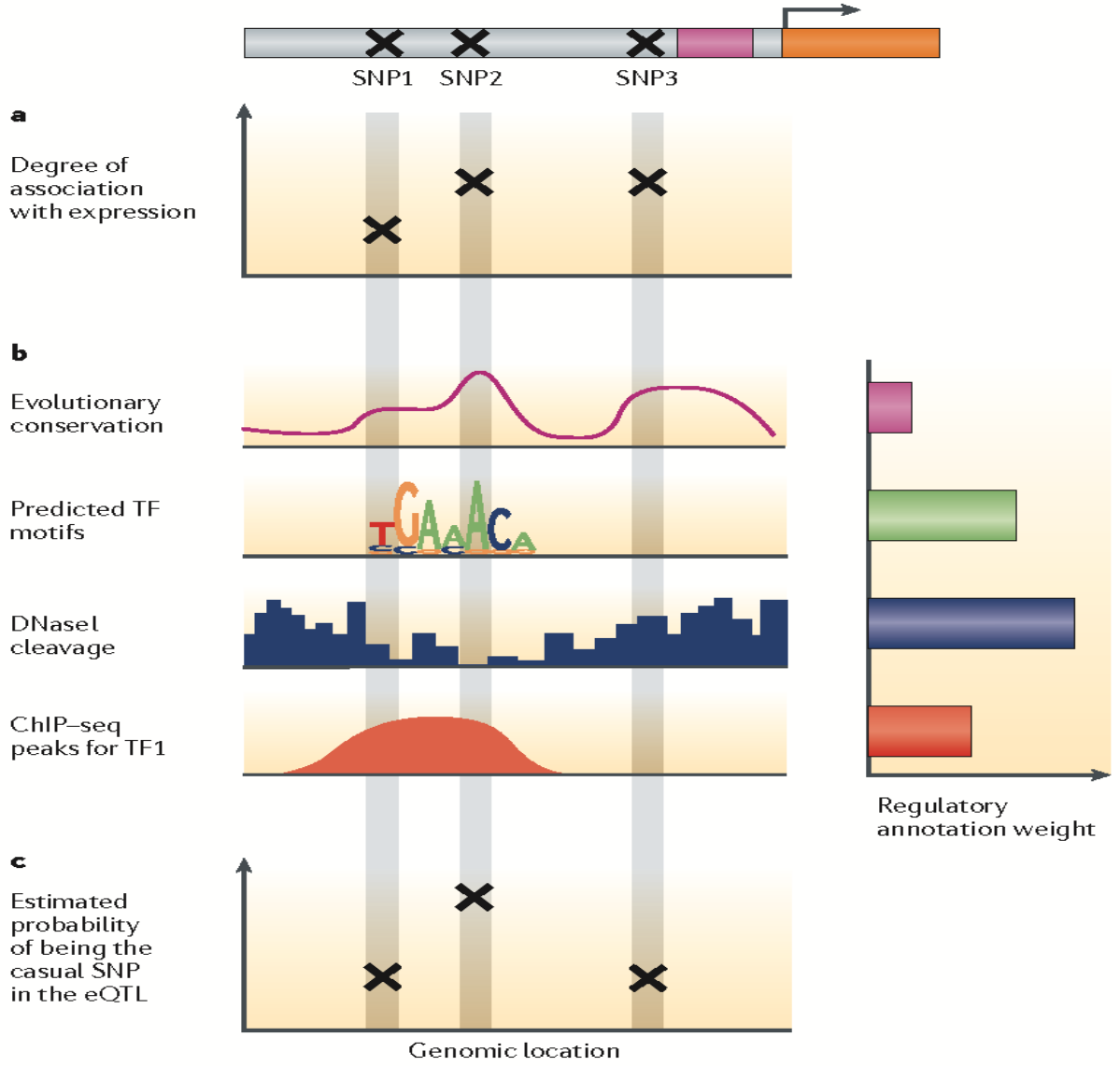
Weingarten-Gabbay and Segal (2014) Hum Genet

Layer 4: From general to specific: individual variation in sequence elements and its impact on quantitative phenotypes and disease

- Catalog and analysis tools can aid investigators in their selection of likely functional variants

- Experimental and computational technology development can hasten progress towards the realization of necessary high-throughput and robust tools.

Incorporating conservation and regulatory annotations to prioritize SNVs



Levo and Segal (2014) NATURE REVIEWS | **GENETICS**

Functional genomics: Imminent challenges and the role of ENCODE

- ENCODE is positioned to make an enabling contribution
- High-throughput approaches for mapping genomic features (biochemical and otherwise) will be complemented by new tools for high-throughput genome engineering and systematic functional perturbation,
- Focus on areas where the coordinated action of a consortium and large-scale data generation can have the most impact.